



Les dictionnaires électroniques du modèle des classes d'objets

Pierre-André Buvet, Aude Grezka

► To cite this version:

Pierre-André Buvet, Aude Grezka. Les dictionnaires électroniques du modèle des classes d'objets. BLANCO X. et P.-A. BUVET. Les représentations des structures prédicat-arguments, 176 (4/2009), Langages, pp. 63-79, 2009, Langages. halshs-00685941

HAL Id: halshs-00685941

<https://shs.hal.science/halshs-00685941>

Submitted on 6 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LES DICTIONNAIRES ELECTRONIQUES DU MODELE DES CLASSES D'OBJETS

Pierre-André Buvet
Aude Grezka
laboratoire Lexiques Dictionnaires Informatique (LDI)
UMR 7187 CNRS-Université Paris 13

L'étude des faits de langue comporte une dimension théorique et une dimension applicative. La première dimension permet d'analyser et d'interpréter les phénomènes étudiés. La deuxième dimension exploite les résultats de la première pour réaliser des produits spécifiques. Les applications possibles sont diversifiées : didactique des langues, thérapeutique, traitement de l'information, etc. La primauté de la dimension théorique sur la dimension applicative n'empêche pas des interactions entre les deux dimensions : une théorie doit être révisée lorsque des défaillances surviennent pendant la mise en place d'une application.

Ce point de vue est illustré par la description des structures prédicat-argument dans des bases de données lexicales. Les particularités des structures et le mode de fonctionnement de leurs constituants sont détaillés et explicités dans un cadre théorique donné. La modélisation sous-jacente à la conception des bases de données est améliorée lorsque sa capacité descriptive est prise en défaut au moment de l'enregistrement des données.

Le modèle des classes d'objets est l'approche théorique des structures prédicat-argument dont nous discutons. Il s'agit d'une théorie descriptive du type lexicaliste, *cf.* GROSS G. 1995a, LE PESANT & MATHIEU-COLAS 1998. A partir d'une définition minimale des constituants phrastiques, on considère les mots comme autant de points d'entrée pour effectuer des descriptions présentées sous forme de taxinomies.

Les descriptions effectuées dans le cadre théorique de référence ont comme objectif applicatif la réalisation de dictionnaires électroniques à large couverture dédiés au traitement automatique des langues, *cf.* GROSS G. 1992, MATHIEU-COLAS 1994. Ces dictionnaires ont comme principale particularité d'associer toutes les entrées lexicales à des informations métalinguistiques standardisées qui sont suffisamment explicites pour donner lieu à des procédures informatisées. De ce fait, ils apparaissent comme un moyen de formaliser les descriptions effectuées. Nous discutons des interactions entre l'élaboration des dictionnaires et la modélisation linguistique qui sous-tend leur conception.

Nous présentons le point de départ du modèle des classes d'objets et la structuration des dictionnaires électroniques qui en a découlé. Nous discutons ensuite des difficultés d'application de la théorie du point de vue de l'encodage des prédicats dans les premières versions des dictionnaires et des solutions théoriques qui ont permis de les dépasser. Enfin,

nous montrons comment ces solutions sont intégrées dans les nouvelles versions des dictionnaires électroniques.

1. La notion de classe d'objets

Nous évoquons les observations des faits de langue qui justifient les programmes de recherche sur les classes d'objets. Nous rappelons les principales particularités du cadre théorique. Nous exposons le travail lexicographique qui s'en est suivi.

1.1. Constat phénoménologique

D'un point de vue lexical, les langues ont comme caractéristiques essentielles la polysémie, la polymorphie, le figement, la paraphrase et la vraisemblance d'occurrence.

La polysémie est définie comme un type d'ambiguïté d'ordre sémantique¹. Par exemple, le nom *main* a plusieurs acceptions : <partie du corps> dans *Ils se serrent la main* ; <humain_collectif> (domaine : marine) dans *Une main fera des heures supplémentaires* ; <inanimé_concret_collectif> (domaine : papeterie) dans *Le papier se vend à la main ou à la rame* ; etc.

La polymorphie est une caractéristique des mots lexicaux qui partagent une même racine tout en étant sémantiquement apparentés. Ils relèvent soit de catégories grammaticales différentes – *développer* et *développement* – soit d'une même catégorie grammaticale – *développement*, *développeur*. Dans le second cas de figure, la variation porte sur le mode de construction des mots : en l'occurrence, la suffixation en *-ement* et la suffixation en *-eur*.

La polylexicalité est la première condition du figement, cf. MEJRI 1997 et 2003. Le caractère figé d'une séquence est mis en évidence par des critères de nature sémantique tels l'opacité dans *dur à la détente* – la valeur de la suite n'est déductible ni de celle de l'adjectif ni de celle du nom – ou la non commutabilité dans *porte-flingue* – (*porte* + **soutient*)-*flingue*/*porte*-(*flingue* + **fusil*) – et de nature syntaxique tels l'absence d'insertion dans *blond vénitien* – *blond* (*E* + **très*) *vénitien* – ou l'impossibilité d'un effacement dans *inférieur à la moyenne* – *inférieur* (**E* + à la moyenne). Ils permettent également d'établir l'importance du figement : le degré de figement d'une séquence dépend du nombre de critères qui la caractérisent, cf. GROSS G. 1996.

La paraphrase est la manifestation du concept de performance, cf. CHOMSKY 1975. Les diverses façons de reformuler un énoncé sont notamment fondées sur la synonymie, cf. MEL'CUK 1988. Les trois paires de synonymes *bourrin-cheval*, *débouler-dévaler* et *piste-terrain* permettent d'exprimer la phrase suivante *Le cheval déboule sur la piste* de sept autres manières : *Le bourrin déboule sur la piste*, *Le bourrin dévale sur le terrain*, etc.

La vraisemblance d'occurrence est une caractéristique des mots en rapport avec leurs contraintes distributionnelles. Toutes les relations entre les mots ne sont pas équipotentes puisque certaines sont plus congruentes que d'autres, cf. HARRIS 1988. Contrairement au verbe *regarder*, le verbe *déguster* n'accepte pas toutes sortes de noms en position complément :

¹ Les autres types d'ambiguïté sont d'ordre morphosyntaxique (*plus* est un adverbe dans *Je n'ai plus d'espoir* ou un verbe au participe passé dans *Ils se sont plus*), d'ordre syntaxique (*Luc regarde manger des poulets* a deux interprétations possibles), etc..

Luc regarde (un avion + une chaise + un gâteau + un manteau + une pomme + un radiateur)

*Luc déguste (*un avion + *une chaise + un gâteau + *un manteau + une pomme + *un radiateur)*

La vraisemblance d'occurrence de *gâteau* ou *pomme* à droite de *déguster* est analysée comme une relation appropriée entre le verbe et les deux noms.

Le modèle des classes d'objets est conçu comme un dispositif théorique qui permet de traiter la polysémie, la synonymie et le figement à partir des relations d'appropriation.

1.2 Remarques méthodologiques

Après avoir rappelé le postulat et les principes d'analyse constitutifs de la théorie, nous montrons l'intérêt des classes d'objets pour décrire les unités lexicales. Nous précisons ensuite les particularités de l'analyse du figement.

Le modèle des classes d'objets postule que toute phrase simple se rapporte à une structure prédicat-argument telle que le prédicat est un verbe, *lire* dans *Luc lit un journal*, un nom, *câlin* dans *Luc fait un câlin à Léa*, un adjectif, *fatigant* dans *Luc est fatigant*, ou une préposition, *sous* dans *Le livre est sous la table*. Les arguments occupent les positions sujet ou complément sous la forme d'un groupe nominal, *Luc dresse un caniche*, d'une complétive, *Luc dit que Léa viendra*, ou d'une infinitive, *Luc pense partir*. Des représentations fonctionnelles sont associées aux structures prédicat-argument correspondant aux phrases simples : **lire (Luc, journal)**, **câlin (Luc, Léa)**, **fatigant (Luc)**, **sous (livre, table)**, cf. GROSS M. 1981. Un argument de nature phrastique donne lieu à une représentation fonctionnelle qui s'incorpore dans une autre représentation fonctionnelle : **dire (Luc, venir (Léa))**.

Les principes d'analyse sont des règles méthodologiques qui permettent d'effectuer la description des faits de langue à partir du postulat fondamental de la théorie. Ils sont nécessaires à l'élaboration de taxonomies. Celles-ci doivent être suffisamment cohérentes et exhaustives pour dégager des régularités qui permettent d'expliquer le fonctionnement des unités linguistiques, cf. GROSS M. 1976. Nous présentons rapidement les principes les plus fondamentaux.

Le premier principe porte sur la nature nécessairement linguistique des descripteurs métalinguistiques utilisés. Autrement dit, une langue donnée se présente comme un système qu'il convient d'étudier indépendamment de facteurs extralinguistiques, cf. HARRIS 1976. Le second principe stipule que la phrase simple est le cadre d'analyse des unités lexicales. C'est une conséquence du premier principe puisqu'il s'agit de préciser la nature des propriétés linguistiques utilisées comme descripteurs. Le troisième principe précise que l'étude du lexique s'effectue conjointement sur les plans morphologique, syntaxique et sémantique de telle sorte que le sens des mots est défini en fonction de leurs caractéristiques morphosyntaxiques. Le troisième principe est un corollaire du deuxième car c'est le contexte phrastique qui rend possible l'analyse morphologique, syntaxique et sémantique du lexique.

La notion de classe d'objets résulte du dispositif théorique fondé sur le postulat et les principes d'analyse. Une classe d'objets est définie : (i) en compréhension par des propriétés

linguistiques, qui sont essentiellement de nature configurationnelle et combinatoire ; (ii) en extension par l'ensemble des items dotés de ces propriétés, cf. GROSS G. 1994. Par exemple, des substantifs comme *camembert*, *frite*, *steak haché* sont caractérisés par la classe d'objets <aliment> parce qu'ils sont tous en position complément des verbes *avalier*, *engloutir*, *ingurgiter* et *manger* qui définissent cette classe. La relation d'appropriation entre les verbes et les noms est fondamentale pour constituer la classe <aliment> : tout substantif qui se combine conjointement avec les quatre verbes est un nom d'<aliment>.

Les variations d'occurrence entre les mots constituent un outil pour la description sémantique : les mots qui partagent les mêmes restrictions de sélection tendent à former des ensembles sémantiquement cohérents, les classes d'objets

Les classes d'objets permettent de rendre compte de la polysémie et de la synonymie, cf. GROSS G. & CLASS 1997 et GROSS G. 1998. En ce qui concerne les arguments, un nom polysémique comme *artère* a une double caractérisation sémantique selon qu'il se combine avec des prédicats appropriés de la classe <voie> comme *rouler sur* ou avec des prédicats appropriés à la classe <vaisseau sanguin> comme *couler dans*. Pour ce qui est des prédicats, un adjectif polysémique comme *abandonné* est rapporté à deux emplois distincts selon qu'il se combine avec : (i) des noms de <membre familial>, *Une femme abandonnée* ; (ii) des noms de <bâtiment>, *Un immeuble abandonné*, ou d'<agglomération>, *Un village abandonné*. Le traitement de la synonymie procède de celui la polysémie prédicative : la polysémie de l'adjectif *abandonné* est rapportée aux synonymes de chacun des deux emplois : (i) *Une femme (abandonnée + délaissée + *vide)* ; (ii) *Un (immeuble +village) (abandonné + *délaissé + vide)*.

Les vraisemblances d'occurrence participent à l'analyse du figement ; elles sont maximales lorsque les séquences sont totalement figées : *pomme de (terre + *sol + *terrain + *glaise)/ (pomme + *poire + *fruit) de terre*. Par ailleurs, le modèle des classes d'objets définit les mots comme des unités monolexicales (*balayeur*) ou des unités polylexicales (*agent d'entretien*) et rapporte le fonctionnement des unités à forme complexe à celui des unités à formes simples. Leur spécificité morphologique mise à part, le verbe simple *fuir* et la locution verbale *prendre la poudre d'escampette* sont analysés de la même façon ; ils sont synonymes, MATHIEU-COLAS 1995.

Le modèle des classes d'objets s'est d'emblée défini comme une théorie dédiée au traitement automatique des langues. Le principal apport applicatif est la conception et l'élaboration de dictionnaires implémentables dans les systèmes qui manipulent de l'information textuelle.

1.3 Lexicographie informatique

Nous évoquons les caractéristiques des premières versions des dictionnaires électroniques fondés sur la notion de classe d'objets. Il y a deux catégories de dictionnaires syntactico-sémantiques : les dictionnaires des arguments et les dictionnaires des prédicats.

La macrostructure des dictionnaires des arguments est constituée de noms élémentaires, c'est-à-dire des substantifs qui sont incompatibles avec la fonction prédicative. La microstructure comporte l'entrée, quatre sortes de descripteurs métalinguistiques (la catégorie grammaticale et une spécification de genre/le trait syntactico-sémantique/la classe d'objet/le domaine

d'emploi) et diverses traductions de l'entrée en rapport avec l'emploi décrit². Les dictionnaires des arguments décrivent aussi bien les noms simples que les noms composés. Par contre, ils sont distingués selon qu'ils se rapportent à des noms communs ou à des noms propres. Un dictionnaire d'expressions figées complète la description lexicographique des arguments ; il décrit les locutions comportant un substantif du dictionnaire des noms communs. Le tableau 1 montre comment des noms de <voie> sont décrits dans les dictionnaires des arguments, cf. MATHIEU-COLAS 1998.

type dictionnaire	champ entrée	champ morpho-syntaxique	champ syntactico-sémantique trait	champ syntactico-sémantique classe d'objets	champ sémantique domaine	traduction anglaise
noms communs	<i>chemin</i>	G:nf	T:loc	C:voie	D:transp	E:lane
noms communs	<i>autoroute à quatre voies</i>	G:nf	T:loc	C:voie	D:transp	E:motorway; highway
noms propres	<i>Francilienne</i>	G:nf	T:loc	C:voie	D:transp	E:Francilienne
noms propres	<i>Autoroute du Soleil</i>	G:nf	T:loc	C:voie	D:transp	E:Francilienne
expressions figées	<i>rebrousser chemin</i>					E:to turn back
expressions figées	<i>tous les chemins mènent à Rome</i>					E:all roads lead to Rome

tableau 1

Les dictionnaires des prédicats sont distingués selon la catégorie grammaticale des unités lexicales qui constituent leur macrostructure : le dictionnaire des verbes, le dictionnaire des adjectifs et le dictionnaire des noms. La microstructure des dictionnaires des prédicats verbaux et des prédicats adjectivaux est constituée de l'entrée, d'indications sur sa structure argumentale et de diverses traductions de l'entrée.

dictionnaire des verbes prédictifs

champ entrée	champ premier argument	champ deuxième argument	traduction anglaise
<i>grimper</i>	N0:N<voie>	N1:	E:to climb
<i>arriver par</i>	N0:hum	N1:loc<voie>	E:to arrive by

dictionnaire des adjectifs prédictifs

champ entrée	champ premier argument	champ deuxième argument	traduction anglaise
<i>à pic</i>	N0:N<voie>	D:transp	E:precipitous
<i>asphalté</i>	N0:loc<voie-t>	D:transp	E:tarred

La microstructure du dictionnaire des prédicats nominaux comporte deux descripteurs supplémentaires : une spécification de trait syntactico-sémantique et une spécification de verbe support. Les deux descripteurs sont en rapport étroit dans la mesure où le trait syntactico-sémantique est déterminé par la nature du support du substantif prédictif³.

dictionnaire des noms prédictifs

champ entrée	champ trait	champ premier argument	champ deuxième argument	champ verbe support	traduction anglaise
--------------	-------------	------------------------	-------------------------	---------------------	---------------------

² Dans les exemples, seules les traductions en anglais sont spécifiées.

³ Les verbes supports *avoir*, *éprouver* et *ressentir* sont caractéristiques des noms d'<état>, les verbes supports *effectuer*, *faire* et *procéder* à des noms d'<action> et le verbe support *il y a* <avoir> et *avoir lieu* des noms d'<événement>, cf. GROSS G. 1999.

	syntactico-sémantique				
<i>colère contre</i>	T :éta	N0:hum	N1 :hum	W:éprouver	E:hanger
<i>asphaltage de</i>	T :act	N0:hum	N1:loc<voie>	W:procéder à	E:asphalting
<i>Explosion</i>	T :évé	N0:	N1	W :il y <avoir>	E explosion:

Il s'est révélé à l'usage que la structuration des dictionnaires de prédicats ne permettait pas, en l'état, de faire une description satisfaisante des adjectifs, des noms et des verbes prédicatifs. La révision lexicographique a nécessité de nouveaux développements théoriques.

2 Les prédicats du premier ordre

La mise en œuvre de la première version des dictionnaires électroniques a permis de constater que la théorie ne tenait pas compte de toutes les particularités des prédicats. Aussi, la modélisation s'est enrichie de nouvelles notions pour améliorer leur description. Il s'en est suivi une différence de traitement plus marquée entre les prédicats et les arguments.

2.1 La spécificité des prédicats

Le projet lexicographique fondé sur le modèle des classes d'objets a rencontré des difficultés pour décrire les prédicats parce que leurs particularités n'étaient pas suffisamment prises en compte.

Les noms élémentaires sont catégorisés par des traits généraux : (i) inanimé concret, *galet* dans *Luc lance un galet dans la mer* ; (ii) inanimé abstrait, *mètre* dans *La planche mesure deux mètres* ; (iii) locatif, *maison* dans *Luc habite dans cette maison* ; (iv) végétal, *herbe* dans *Luc ramasse de l'herbe pour son lapin* ; (v) animal, *lapin* dans *Luc donne de l'herbe à son lapin*⁴. Les classes d'arguments constituent une subdivision sémantique plus fine des noms élémentaires. Leur définition est fondée sur des critères syntaxiques. Par exemple, la classe <vêtement> correspond aux noms élémentaires qui ont comme prédicats appropriés définitionnels les verbes *enfiler*, *essayer*, *mettre* et *porter* : *Luc (enfile + essaye + met + porte) une (casquette + chemise + culotte + gabardine + tunique...)*.

Les classes d'arguments contribuent à l'analyse de la polysémie des prédicats. Les différentes acceptions d'un mot à caractère prédicatif sont rapportées aux particularités de sa structure argumentale. Les arguments sont spécifiés par des traits généraux et, lorsque les traits ne sont pas assez discriminants, par des classes d'objets, voire par une ou plusieurs unités lexicales. Par ailleurs, chaque acception est caractérisée par une classe sémantique.

adjectif	indice	classe sémantique	N0
<i>cotonneux</i>	1	<apparence>	<i>nuage</i>
<i>cotonneux</i>	2	<couverture naturelle>	<végétal>
<i>cotonneux</i>	3	<aspect météorologique>	<i>ciel</i>
<i>cotonneux</i>	4	<fatigue>	<humain>
<i>cotonneux</i>	5	<qualité stylistique>	<manière d'écrire>
<i>cotonneux</i>	6	<consistance>	<fruit>
<i>cotonneux</i>	7	<qualité sonore >	<bruit>

⁴ Le trait humain caractérise à la fois des noms élémentaires (*homme*) et des noms prédicatifs (*fil*s). Ils constituent une catégorie syntactico-sémantique à part, cf. GROSS. G. 1995b.

Le traitement de la polysémie permet de catégoriser sémantiquement les prédicats sous forme de classes d'objets. Par exemple, la classe <consistance> caractérise une des significations de *cotonneux*, *charnu*, *crémeux épais*, *ferme*, *glaireux*, etc.

Les relations d'appropriation ont un rôle central dans la définition des classes d'arguments. Elles participent à la définition des classes de prédicats lorsque ces dernières procèdent de l'analyse des structures argumentales des prédicats polysémiques. Cependant, les propriétés syntactico-sémantiques des arguments ne permettent pas de désambiguïser l'ensemble des prédicats. Les deux acceptions de la séquence verbale *voir loin* ('avoir une vue qui porte loin' et 'être prévoyant') ne sont pas déductibles de leur structure argumentale puisque les deux prédicats ont le même domaine d'arguments (un nom d'<humain>) : *Luc voit loin*, GREZKA 2009.

D'autres relations d'appropriation permettent de caractériser sémantiquement les prédicats : des adverbes appropriés comme *distinctement* et *clairement* pour les verbes de <parole> (*Luc s'exprime distinctement*) ; des verbes supports appropriés comme *commettre* pour les noms de <crime-délit> (*Luc a un commis un vol à l'étalage*) ; des adjectifs appropriés comme *percutant* pour les noms de <style> (*Luc a une manière d'écrire percutante*) ; etc. Elles sont néanmoins insuffisantes pour définir toutes les classes de prédicats.

La polymorphie est une propriété des arguments et des prédicats, mais elle est plus contrainte pour les arguments que pour les prédicats. Sur le plan quantitatif, la polymorphie argumentale est moins fréquente que la polymorphie prédictive. Sur le plan morphologique, les arguments polymorphes sont des adjectifs et des noms (*pétrolier* et *pétrole*) tandis que les prédicats polymorphes sont des adjectifs, des noms et des verbes (*jaloux*, *jalousie* et *jalouser*). Sur le plan syntaxique, les arguments adjectivaux sont uniquement du type épithète et obligatoirement en postposition : *Sa production pétrolière*/* *Sa pétrolière production* /**Sa production est pétrolière*⁵. La polymorphie des prédicats n'affecte pas leur syntaxe : *Mon ami est jaloux de toi*/J'ai un ami jaloux de toi ; *Mon ami éprouve de la jalousie envers toi*/La jalousie de mon ami envers toi ; *Stupidement, mon ami te jalouse* /*Te jalouser est stupide de la part de mon ami*. Sur le plan sémantique, l'interprétation des arguments reste constante quelle que soit leur forme : l'argument nominal et l'argument adjectival sont équivalents dans *La production (pétrolière + de pétrole)*. Par contre, selon leur forme, l'interprétation des prédicats peut rester constante ou bien présenter des variations importantes. Le verbe *détruire* et le nom *destruction* donnent lieu à une même interprétation processive du prédicat dans *Luc détruit la maison* et *Luc a procédé à la destruction de la maison*. Les adjectifs *destructible* et *détruit* sont reliés à *détruire* et *destruction* sur le plan morphosémantique, mais l'interprétation des prédicats n'est pas processive : *La maison est (destructible + détruite)*.

La description des arguments est hiérarchisée : les nœuds inférieurs se rapportent aux noms élémentaires ; les nœuds supérieurs correspondent aux traits sémantiques généraux ; les nœuds intermédiaires sont constitués de classes et d'hyperclasses. La hiérarchisation n'est pas une ontologie stricte dans la mesure où elles procèdent de l'analyse ascendante⁶ et autorisent les héritages multiples⁷, cf. LE PESANT 1998.

⁵ Sur les rapports entre les compléments du nom et les adjectifs de relation, cf. BARTNING 1976.

⁶ C'est-à-dire les classes homogènes sont les unités minimales de description des unités linguistiques à partir desquelles il est possible de faire des regroupements à des niveaux supérieurs.

⁷ Les noms de <vin> sont rattachés aux l'hyperclasses <liquide> et <aliment>.

Une hiérarchisation des prédicats calquée sur celle des arguments est ontologiquement satisfaisante, mais linguistiquement non pertinente⁸. Par exemple, le rattachement des adjectifs *en colère* et *colérique*, respectivement, aux hyperclasses <affect> et <comportement> rendrait difficile l'interprétation de leur relation dans *Il se met souvent en colère mais n'est pas vraiment colérique* car les hyperclasses appartiennent à des embranchements différents. Par ailleurs, la polymorphie prédicative interfère avec le double rattachement de certaines de classes de noms prédicatifs aux traits généraux action et événement. Le substantif *assassinat*, comme tous les noms de <crime et délit>, s'interprète à la fois comme un prédicat d'action – *Luc a commis l'assassinat contre Tom* – et comme un prédicat d'événement – *Il y a eu un assassinat*. Par contre, le verbe *assassiner* s'interprète uniquement comme un prédicat d'action – *Luc a assassiné Tom*. Autrement dit, le nom et le verbe, malgré leur unité prédicative, ont des propriétés linguistiques spécifiques qui sont difficilement compatibles avec une représentation hiérarchisée. La mise en place d'une taxonomie des prédicats comparable à celle des arguments contrevient au principe d'analyse selon lequel il n'y a pas de métalangue hors de la langue, cf. *supra*.

La spécificité des prédicats apparaît clairement au regard de : (i) l'impossibilité de recourir systématiquement à la relation d'appropriation pour définir des classes sémantiques ; (ii) leur caractère nettement polymorphique ; (iii) la difficulté d'élaborer une taxonomie hiérarchisée. De nouveaux outils théoriques se sont avérés nécessaires pour analyser et décrire les prédicats.

2. 2 Les emplois prédicatifs

Les problèmes rencontrés pendant l'élaboration des dictionnaires électroniques de prédicats ont conduit à préciser la notion d'emploi. Elle a permis de mieux prendre en compte la polymorphie prédicative et d'aboutir à la notion de racine prédicative, cf. BUVET & alii 2006. Cette dernière permet d'expliquer la variété des interprétations qui peut affecter les différentes formes d'un prédicat.

La notion d'emploi est présentée comme fondamentale pour l'analyse des prédicats, cf. GROSS G.1999. Un emploi est défini initialement à partir de huit paramètres : 1) un domaine d'arguments, *occire* a des noms d'humain comme premier et deuxième arguments ; 2) un sens, *occire* a comme synonymes *assassiner* et *tuer* ; 3) une forme, *occire* est un verbe ; 4) une actualisation, la conjugaison de *occire* est défective ; 5) un système aspectuel, *occire* est un duratif perfectif ; 6) des transformations, *occire* a une forme pronominale réfléchie (*il s'est occis*) ; 7) un domaine, *occire* appartient à la langue générale ; 8) un niveau de langue, *occire* appartient au registre soutenu.

Les huit paramètres n'ont pas tous la même importance. Nous négligeons ici le domaine, le niveau de langue et les transformations car ces paramètres n'interagissent pas avec les autres pour définir un emploi. L'hétérogénéité des autres paramètres n'est qu'apparente car ils sont interdépendants. On peut distinguer des interdépendances d'ordre sémantique, d'une part, et d'ordre distributionnel, d'autre part. La première catégorie d'interdépendances concerne le domaine d'argument, le sens, la forme et le système aspectuel ; la deuxième catégorie d'interdépendances a trait au domaine d'arguments, à la forme et à l'actualisation.

⁸ LEVIN 1993 fait une description des prédicats verbaux fondée uniquement sur une ontologie

Un emploi prédicatif est de nature linguistique, il est directement observable. Un prédicat est de nature métalinguistique, on le met en évidence par l'observation des emplois prédicatifs. Autrement dit, un emploi prédicatif est une occurrence d'un prédicat sous forme d'adjectif, de nom ou de verbe, cf. BUVET à paraître (1).

La polymorphie s'explique par les interactions entre la forme et le sens. Un seul emploi prédicatif peut correspondre à un prédicat (l'adjectif *hagard*). Lorsque plusieurs emplois se rapportent à un même prédicat, ils peuvent avoir la même forme (les adjectifs *flabellé* et *flabelliforme*) ou des formes différentes (le verbe *croire*, le nom *croissance* et l'adjectif *croissant*). Un prédicat est conçu comme une racine prédicative : celle-ci est un lemme lorsque le prédicat a un seul emploi (**flappi**), une base lorsque le prédicat correspond à plusieurs emplois de forme différente (**cour-** est la racine prédicative des emplois *courir* et *course*).

La polysémie peut s'analyser au moyen des interactions entre le domaine d'arguments et le sens. La racine **boucl-** est commune à plusieurs prédicats ; chaque prédicat a un domaine d'arguments spécifique : *Luc boucle ses cheveux* +Léa/*Luc boucle ce numéro de journal*. La polymorphie participe à la description de la polysémie car, d'une manière générale, les emplois prédicatifs présentent des variations morphologiques en fonction du sens des prédicats : *Luc fait des bouclettes* à (ses cheveux +Léa) /*Luc procède au bouclage de ce numéro de journal*. Les domaines d'arguments ne sont pas toujours suffisants pour caractériser un prédicat. La racine **grog-** se rapporte à deux prédicats distincts qui partagent le même domaine d'arguments : *Luc (grogne + émet un grognement)* / *Luc (grogne + est grognon)*.

La spécification de la classe sémantique est fondamentale pour caractériser les prédicats. Une classe sémantique est fondée sur la synonymie ou bien l'hyperonymie. La classe <colère> caractérise les prédicats **colèr-**, **courrou-**, **irrit-**, **rag-**, **rogne**, etc ; elle est définie à partir du prédicat le plus représentatif dont la forme nominale est l'éponyme de la classe, cf. BUVET 2008. La classe <mode de nourriture> est constituée des prédicats hyponymes **anthropophag-**, **autophagi**, **cannibale**, **carnassier**, **carnivore**, **carpophag-**, etc. dont l'hyponyme est substantif éponyme de la classe.

En revanche, elle est nécessaire mais non suffisante pour caractériser les emplois prédicatifs. Ils peuvent avoir des interprétations différentes lorsqu'ils héritent de la classe sémantique de leur prédicat commun. Le prédicat **gastrectomie** est rattaché à la classe <acte médical> ; il correspond à deux emplois différents : l'un est de nature processive, *Luc fait une gastrectomie à Tom* ; l'autre est de nature événementielle, *La gastrectomie a eu lieu à 9 heures*.

La caractérisation sémantique d'un prédicat se limite à la spécification d'une classe, celle d'un emploi doit tenir compte également de son trait et de son aspect inhérent, cf. *infra*. Par exemple, les traits 'action' et 'événement' permettent de distinguer les deux emplois de **gastrectomie**. De plus, il arrive que des emplois s'interprètent différemment malgré leur trait commun. Le verbe *lire* équivaut à deux emplois verbaux du prédicat **I(ir-/ect-)** : *Luc lit un livre* ; *Luc lit surtout des romans*. Les deux emplois partagent le trait 'action' mais chacun a un aspect inhérent spécifique : respectivement, le duratif perfectif et le duratif imperfectif⁹.

Les interdépendances d'ordre distributionnel permettent de spécifier la construction de chaque emploi prédicatif et de préciser, par rapport à cette construction, la nature des arguments du

⁹ Différents tests permettent d'établir le trait et l'aspect d'un emploi, cf. BUVET & GREZKA 2007, JAYEZ 1996.

prédicat sous-jacent. Une construction résulte de la forme et de l'actualisation d'un emploi prédicatif : un adjectif prédicatif, un nom prédicatif et un verbe prédicatif ont chacun un mode d'actualisation qui leur est propre. Elle procède également du domaine d'arguments du prédicat qui permet de faire état du mode de structuration des arguments par rapport à l'emploi prédicatif, du nombre d'arguments spécifiés et de leur caractère obligatoire ou facultatif en position complément.

La structure argumentale des emplois prédicatifs ne correspond pas nécessairement au domaine d'arguments de leurs prédicats. La racine **hachur-** représente un prédicat dyadique dont la forme adjectivale est *hachuré*, la forme nominale est *hachure* et la forme verbale est *hachurer* : *Luc hachure la feuille/Luc fait des hachures sur la feuille/La feuille est hachurée/La feuille a des hachures*. L'emploi verbal et le premier emploi nominal exigent que les deux arguments soient stipulés, l'emploi adjectival et le deuxième emploi nominal autorisent uniquement la stipulation d'un argument. De plus, il y a une variation diathésique entre les deux premiers emplois et les deux derniers : l'argument en position complément passe en position sujet. L'emploi verbal et le premier emploi nominal ont une interprétation processive, les deux autres emplois s'interprètent comme une propriété résultative. Chaque emploi a une construction en rapport avec sa forme.

La spécification des particularités morphosyntaxiques et syntactico-sémantiques des arguments permet de mieux définir les emplois. Par exemple, il y a deux emplois du verbe *aimer* l'un se rapporte à un prédicat d'<amour> (*Luc aime Léa*) , l'autre à un prédicat de <goût> (*Luc aime rêver*) ; les deux emplois sont caractérisés par la construction **X0 V X1** ; la position **X0** est occupée dans les deux cas par un groupe nominal constitué d'un nom d'<humain> ; la position **X1** est occupée uniquement par un groupe nominal constitué d'un nom d'<humain> dans le premier cas, par un groupe nominal, une complétive ou une infinitive dans le deuxième cas.

Les critères définitionnels des emplois prédicatifs que nous proposons sont au nombre de sept : 1) la racine prédictive ; 2) la classe sémantique ; 3) le type sémantique ; 4) l'aspect inhérent ; 5) la construction ; 6) les particularités morphosyntaxiques des arguments ; 7) leurs particularités syntactico-sémantiques. Il s'agit d'autant de paramètres d'analyse dont il faut tenir compte pour l'encodage des emplois prédicatifs dans les dictionnaires.

Avant de discuter de la nouvelle microstructure du dictionnaire des prédicats, nous présentons trois outils théoriques développés pour l'analyse des emplois prédicatifs.

2.3 Traitement des emplois prédicatifs

La mise en œuvre de la notion d'emploi prédicatif a fait apparaître l'importance de l'actualisation pour l'analyse des adjectifs, des noms et des verbes prédicatifs. Il y a trois oppositions majeures qui permettent d'améliorer la description des emplois prédicatifs.

Actualisation prédictive vs actualisation argumentale

Les actualisateurs sont distingués selon qu'ils se rapportent à des emplois prédicatifs ou bien à des arguments. Dans le premier cas de figure, les occurrences des actualisateurs dépendent uniquement de la nature des emplois prédicatifs. Le nom prédicatif *gifle* exige *faire* comme verbe support alors que *gifle* exige *donner* comme verbe support : *Luc (donne + *fait) une gifle/Luc (*donne + fait) une bêtise*. Dans le second cas de figure, les occurrences des

actualisateurs procèdent des structures prédicat-argument. En général, le nom *conte* se combine aussi bien avec *un* que *le* quand il est en position complément : *Luc lit (un + le) conte*). Cependant, le verbe *inventer* tend à imposer l'article indéfini à ce substantif : *Luc est en train d'inventer (un + ?le) conte*.

Les actualisateurs ont un mode de fonctionnement et une signification qui les distinguent des autres constituants phrastiques. Ce sont principalement des mots ou de morphèmes qui sont nécessaires à la grammaticalisation des phrases. Ils apportent généralement des informations spécifiques. Elles sont de deux sortes : (i) les informations obligatoires ; (ii) les informations facultatives. Les actualisateurs standard donnent lieu uniquement aux informations du type (i) ; ils sont porteurs de significations indispensables à l'instanciation des prédicats et des arguments dans des phrases grammaticales (par exemple, le temps ou le nombre). Les actualisateurs non standard donnent lieu à des informations des types (i) et (ii). Ils sont souvent porteurs des mêmes significations que les actualisateurs standard et, conjointement, de significations qui se surajoutent à celles-ci (par exemple, l'aspect). Quand les actualisateurs non standard véhiculent uniquement des valeurs surajoutées, ils sont obligatoirement associés à des actualisateurs standard.

Signification lexicale vs signification grammaticale

La signification lexicale subsume les valeurs des prédicats et des arguments tandis que la signification grammaticale subsume les valeurs des actualisateurs, cf. BLANCO & BUVET 2004 et BLANCO dans ce numéro. La signification lexicale est représentée par des classes sémantiques.

La signification grammaticale rassemble des valeurs hétéroclites. Il y a les valeurs de l'actualisation standard (par exemple un type de temps) et les valeurs de l'actualisation non standard (par exemple un type d'intensité). Les valeurs de l'actualisation standard sont celles qu'implique le procédé de grammaticalisation. Le verbe *avait* dans *Luc avait faim* est un verbe support qui permet de conjuguer le nom prédicatif *faim*, cf. GROSS G. 2004.

Les valeurs de l'actualisation non standard ont affaire, d'une part, avec la spécification d'une quantité ou d'une appréciation et, d'autre part, avec la modalisation des prédicats. Pour ce qui est du second point, les valeurs permettent notamment de préciser la façon dont se déroule un processus et d'indiquer la gradation d'un état. Le verbe complexe *mourir de* est une variante de *avoir* dans *Luc mourrait de faim*. Il permet conjuguer le nom prédicatif *faim* tout en indiquant l'intensité forte.

Aspect inhérent vs aspect cotextuel

D'une manière générale, l'aspect se conçoit comme une marque de temps aréférentiel. Les indications du type présent, passé ou futur s'interprètent toujours par rapport à un repère temporel, explicite ou implicite. L'interprétation des informations aspectuelles ne nécessite aucun repère temporel. Elle s'effectue à des niveaux d'analyse distincts selon que l'aspect est inhérent (lexical) ou cotextuel (grammatical). L'aspect inhérent est un critère définitionnel d'un emploi prédicatif, cf. *supra*. L'aspect cotextuel est pris en charge par des actualisateurs. Les actualisateurs aspectuels d'un emploi prédicatif dépendent de son aspect inhérent.

Il y a différents types d'aspect inhérent selon que l'on les emplois prédicatifs sont du type action et événement, d'une part, du type état, d'autre part. Les verbes prédicatifs *nettoyer* (*Luc*

nettoie une assiette) et *jaillir* (*Le renard jaillit de son terrier*) ont respectivement comme particularité aspectuelle le duratif perfectif et le ponctuel.

Les actualisateurs aspectuels relèvent de l'actualisation non standard. Leur association avec les prédicats constitue des indications sur la fréquence des prédicats et leur stade d'avancement (le début, la fin, la continuité d'un processus ou d'un état, etc.). La nature des indications explique les différents types d'aspect cotextuel : le ponctuel, l'itératif, l'inchoatif, le terminatif, le continuatif, etc.

L'aspect inhérent d'un emploi prédicatif explique sa combinatoire avec des actualisateurs aspectuels. Les emplois prédicatifs du type action ou événement dont l'aspect inhérent est le duratif (perfectif ou imperfectif) sont compatibles avec des actualisateurs aspectuels du type inchoatif : *Luc entame une traversée de l'Atlantique/Luc commence à lofer/Il y a eu un début de panique dans l'aéroport*. Ces actualisateurs sont incompatibles avec les prédicats du type action ou événement dont l'aspect inhérent est le ponctuel : **Luc a entamé un clin d'œil/*La vapeur commence à fuser du tuyau d'échappement/*Il y a eu un début d'explosion dans l'aéroport*.

L'aspect intrinsèque des emplois prédicatifs d'état est : (i) le provisoire, *La porte est ouverte* ; (ii) le permanent, *Il est intelligent*. L'adverbe *régulièrement* se combine uniquement avec les emplois prédicatifs du type état dont l'aspect inhérent est le provisoire : *La porte est régulièrement ouverte*. Il est incompatible avec les emplois prédicatifs du type état dont l'aspect inhérent est le permanent : **Luc est régulièrement intelligent*.

3. Les dictionnaires électroniques syntactico-sémantiques

Il y a trois formes de dictionnaires électroniques syntactico-sémantiques PRED-DIC, ARGU-DIC et ACTU-DIC. Le format PRED-DIC concerne les prédicats, le format ARGU-DIC les arguments et le format ACTU-DIC les actualisateurs¹⁰. Nous montrons que la conception du dictionnaire des prédicats résulte de la définition des emplois prédicatifs.

La macrostructure des dictionnaires du type PRED-DIC est constituée des différents emplois prédicatifs. La microstructure des dictionnaires du type PRED-DIC est constituée d'une entrée et de deux catégories de descripteurs : les descripteurs de définition ; les descripteurs de conditions. Les descripteurs n'ont pas la même fonction selon leur catégorie. La fonction des descripteurs de définition est d'indiquer les propriétés sémantiques des emplois. La fonction des descripteurs de conditions est de vérifier que les emplois prédicatifs ont les propriétés remarquables qui les caractérisent¹¹.

Les descripteurs de définition donnent lieu à quatre sortes de spécification : la racine prédicative, la classe sémantique, le type sémantique et l'aspect inhérent. Les spécifications de racine et de classe permettent de préciser la nature de l'articulation entre les emplois prédicatifs et les prédicats. Les interprétations des emplois prédicatifs résultent des quatre spécifications.

¹⁰ Le format définitif du dictionnaire ACTU-DIC n'est pas encore établi.

¹¹ Les deux catégories de descripteurs sont constituées de descripteurs formels, c'est-à-dire ils sont normalisés et totalement explicites. La notion de descripteur formel est fondamentale car les informations métalinguistiques sont exploitables sur le plan informatique à condition d'échapper à toute irrégularité et à toute ambiguïté.

entrée (emploi)		Définition			
emploi	i n d i c e	racine prédicative	Classe	type	aspect inhérent
dévor	1	dévor	<alimentation>	Act	duratif perfectif
dévor	2	dévor	<lecture>	Act	duratif perfectif

Les descripteurs de conditions donnent lieu à trois sortes de spécifications : la construction ; la distribution morphosyntaxique ; la distribution syntactico-sémantique. Les spécifications de construction permettent de faire état du mode de structuration des arguments par rapport à l'emploi prédicatif, du nombre d'arguments spécifiés et de leur caractère obligatoire ou facultatif en position complément. Les spécifications de distribution morphosyntaxique portent sur les particularités formelles des arguments. Les descripteurs consistent à préciser quelles constructions occupent les positions argumentales (groupes nominaux, complétives ou infinitives). Les spécifications de distribution syntactico-sémantique portent sur les particularités sémantiques des arguments. Les descripteurs sont des classes ou des hyperclasses sémantiques.

entrée (emploi)		Conditions		
emploi	I n d i c e	Construction	distribution 1	distribution 2
dévor	1	X0 V X1	X0=GN X1=GN	X0=<humain> X1=<aliment>
dévor	2	X0 V X1	X0=GN X1=GN	X0=<humain> X1=<texte>

Les descripteurs de conditions ne désambigüisent pas toujours une forme prédicative donnée. Les propriétés combinatoires des emplois prédicatifs peuvent contribuer à la levée d'ambigüité. Elles sont décrites dans des grammaires locale, cf. GROSS M. 1995, MAUREL 1993, BUVET à paraître (2).

4. Conclusion

L'élaboration de dictionnaires électroniques dédiés au traitement automatique des langues est un contexte applicatif qui permet de valider les options théoriques d'une modélisation linguistique. La structuration des bases de données exige que les faits de langue soient représentés d'une façon formalisée, c'est-à-dire les représentations doivent être totalement explicites et normalisées. Cette exigence de formalisation implique une théorie linguistique qui satisfait aux critères scientifiques de cohésion, de reproductibilité et de vérifiabilité.

Dans le cas du modèle des classes d'objets, la conception initiale des dictionnaires a entraîné une révision du statut théorique des prédicats. Il s'en est suivi la notion d'emploi prédicatif.

La mise en œuvre de cette notion a également montré l'importance de l'actualisation dans la description des adjectifs, des noms et des verbes prédicatifs.

L'encodage des structures prédicat-argument dans des bases de données lexicales a permis de préciser quelles sont les interactions entre les dimensions théorique et applicative d'une étude des faits de langue. Un autre moyen de vérifier la validité d'une modélisation linguistique est de mettre en œuvre un étiqueteur sémantique des phrases fondés sur des ressources dictionnairiques de qualité, cf. BUVET & alii 2007. Ce travail est en cours de réalisation.

Bibliographie

- BARTNING E. 1976, « Remarques sur la syntaxe et la sémantique des pseudo-adjectifs dénominaux en français », *Revue Romane*.13, Institut d'Etudes Romanes de l'Université de Stockholm, Stockholm.
- BLANCO X. et BUVET P.-A. 2004, « Verbes supports et significations grammaticales. Implications pour la traduction espagnol-français », *Lingvisticae Investigationes* 27 (2), John Benjamins B.V., Amsterdam.
- BUVET P.-A. 2008, « Vers l'élaboration d'un dictionnaire unique des prédicats du français : DEESSE. Dictionnaire Électronique Syntactico-Sémantique » *Cahiers du CENTAL* 5, Université de Louvain-la-Neuve, Louvain-la-Neuve.
- BUVET P.-A. à paraître (1), « Des mots aux emplois : la représentation lexicographique des prédicats », in *Le Français moderne*.
- BUVET P.-A. à paraître (2), « Quelles procédures d'étiquetage pour la gestion de l'information textuelle électronique », in *Information grammaticale*.
- BUVET P.-A. et C. GIRARDIN, G. GROSS, C. GROUD 2005, « Les prédicats d'<affect> », *LIDIL*, 32, Presses Universitaires de Grenoble, Grenoble.
- BUVET P.-A., CARTIER E., ISSAC F., MEJRI S. 2007, « Dictionnaires électroniques et étiquetage syntactico-sémantique », *Actes des 14e journées sur le Traitement Automatique des Langues Naturelles*, IRIT Press, Toulouse.
- BUVET P.-A. & GREZKA 2007, « Elaboration d'outils méthodologiques pour décrire les prédicats du français ». *Lingvisticae Investigationes* 30 (2), John Benjamins B.V., Amsterdam.
- CHOMSKY N. 1975, *Aspect de la théorie syntaxique*, Seuil, Paris.
- GREZKA A. 2009, *Polysémie des verbes de perception visuelle*, L'Harmattan, Paris.
- GROSS G. 1992. "Forme d'un dictionnaire électronique". *L'environnement traductionnel, actes du colloque de Mons*, AUPELF-UREF.
- GROSS G. 1994, « Classes d'objets et description des verbes », *Langages* 115, Larousse, Paris.
- GROSS G. 1995a. « Une sémantique nouvelle pour la traduction automatique - les classes d'objets », *La tribune des industries de la langue et de l'information électronique* 17-18-19, Paris.
- GROSS G. 1995b, « A propos de la notion d'humain », *Lingvisticae Investigationes supplementa* 17, John Benjamins B.V., Amsterdam.
- GROSS G. 1996, *Les expressions figées en français : noms composés et autres locutions*, Ophrys, Paris-Gap.
- GROSS G. 1998, "Pour une véritable fonction Synonymie dans un traitement de texte", *Langages* 131, Larousse, Paris.
- GROSS G. 1999a, « La notion d'emploi dans le traitement automatique », *La pensée et la langue*, . Wydawnictwo Naukowe AP. Krakow.
- GROSS G. 1999b, « Elaboration d'un dictionnaire électronique », *Bulletin de la Société de linguistique de Paris*, XCIV:1, Paris.

- GROSS G. 2004, « Pour un Bescherelle des prédicats nominaux » *Les verbes supports : nouvel état des lieux, Lingvisticae Investigationes Fascicule spécial*, 27(2), John Benjamins Publishing Co. Amsterdam/Philadelphia.
- GROSS G. et A. CLASS 1997, "Synonymie, polysémie et classes d'objets". *Meta* 42 (1), Presses de l'Université de Montréal, Montréal.
- GROSS M. 1976, « Préface », in BOONS Jean-Paul, Alain GUILLET et Christian LECLERE, *La structure des phrases simples en français. Constructions intransitives*, Droz, Genève-Paris.
- GROSS M. 1981, « Les bases empiriques de la notion de prédicat sémantique », *Langages*, 63, Larousse, Paris
- GROSS M. 1995, « Une grammaire locale de l'expression des sentiments », *Langue française*, 105, Larousse, Paris, pp. 70-87.
- HARRIS Z. 1988, *Language and Information*, Columbia University Press, New-York (traduction française Ibrahim A, et C. Martinot).
- JAYEZ J. 1996, « Référence et aspectualité. Le problème des verbes dits aspectuels », *Cahiers de Linguistique Française* 18, Paris.
- LEVIN B. 1993, *English Verb Classes and Alternations: A Preliminary Investigation*, The University of Chicago Press, Chicago and London:.
- LE PESANT D. 1998, « Utilisation des propriétés des anaphores dans la définition des relations lexicales », *Langages* 131, Larousse, Paris.
- MAUREL D. 1993, « Reconnaissance automatique d'un groupe nominal prépositionnel. Exemple des adverbes de date », *Lexique*, 11, Presses Universitaires de Lille, Lille.
- MEJRI S. 1997, *Le figement lexical : descriptions linguistiques et structuration sémantique*, Faculté des lettres de la Manouba, Tunis.
- MEJRI S. (ed). 2003, *Polysémie et polylexicalité, Syntaxe et sémantique* 5, Presses de Université de Caen, Caen.
- MATHIEU-COLAS M. 1994, *Les mots à trait d'union. Problèmes de lexicographie informatique*, Didier-Érudition, Paris.
- MATHIEU-COLAS M. 1995, "Les mots composés". in *Les concepts opératoires de l'INaLF*, INaLF, Nancy.
- MATHIEU-COLAS M. 1998, « Illustration d'une classe d'objets : les voies de communication », *Langages* 131, Larousse, Paris.
- MEL'CUK I. 1988, « Lexique et paraphrase dans la théorie sens-texte », *Lexique* 6, Presses universitaires du septentrion, Villeneuve d'Ascq